The Effect of Assumptions on Unknown Parameter Values in Forecasting Reliability
of Meeting Effluent Limits

Troy A. Doby[1], Ishwar Devkota[2], Raymond Smith[3]

[1]U. S. Environmental Protection Agency, NRMRL, Sustainable Technology Division,
Systems Analysis Branch, 26 W. Martin Luther King Dr., Cincinnati, OH 45268,
Phone: (513) 569-7360, email: doby.troy@epa.gov
[2]North Carolina Department of Environmental and Natural Resources, Division of
Water Quality, Consturction Grants and Loans Section, 2728 Capital Blvd., Raleigh,
NC 27604, Phone: (919) 715-6222, email: ishwar.devkota@ncmail.net
[3]U. S. Environmental Protection Agency, NRMRL, Sustainable Technology Division,
Systems Analysis Branch, 26 W. Martin Luther King Dr., Cincinnati, OH 45268,
Phone: (513) 569-7161, email: smith.raymond@epa.gov

Abstract

Forecasting tools exist for regulatory agencies to predict the reliability in
meeting effluent limits or total maximum daily loads (TMDLs) from point sources
discharging to water bodies covered under section 303(d) of the Clean Water Act.
Performing 2-D Monte Carlo simulations with these tools requires inputs of known
(e.g., flow, temperature, influent concentrations) and unknown parameter values (e.g.,
autotrophic and heterotrophic growth and decay rates) and distributions. This paper
examines the effects of assumptions about distributions, coefficients of variation, and
parameter correlations on reliability results with several examples.

Introduction
        In previous work it was shown that while design of a wastewater treatment
plant for higher percentiles of flow results in greater expense, it does not necessarily
result in greater reliability in terms of meeting an ammonia or a total nitrogen effluent
concentration limit (Doby et al. submitted). A set of assumptions was made about the
distributions of the loads (e.g., flow, temperature, influent concentrations) and
parameters (e.g., growth and decay rates) when reaching this conclusion. In addition,
an assumption was made about the correlation of the different loading values. The
purposes of the present work are: (1) to demonstrate the effect of assumptions about
different coefficients of variability; (2) to demonstrate the effect of assumptions about
different distributions of parameters; (3) to demonstrate the effect of assumptions
about correlation of input loadings; and most importantly (4) to gauge the robustness
of the previous conclusion for different sets of assumptions for the distributions and
the correlation.

Methods

        In previous work (Doby et al. submitted), the steady state Water Research
Council model (WRC 1984) was used in Monte Carlo simulations to determine
reliability in meeting effluent concentrations . The influent loadings were assumed to

conform to an empirical Pearson Type IV distribution and the flows and temperatures were repeated to provide 20-years of data. In addition, the loadings were assumed to be correlated with one another in a piecewise manner.

In the current work, Crystal Ball 5.0 is used to fit the data for the loadings for both distribution types and correlations. The assumed distribution types and their statistics are contained in Table a. Also contained in Table a are the parameters and their mean values. Monte Carlo simulations were performed with the parameters having Gaussian, lognormal, and uniform distributions. It should be emphasized that neither the mean nor the distributions of these parameters is known *a priori*. In addition, the correlations (if any) of the parameters are unknown. The parameter correlations are, however, not considered in this current work. The effects of different coefficients of variation values, specifically, 0.01, 0.05, 0.10, 0.20, 0.25, and 0.50, are also shown. The standard deviation for the Gaussian and lognormal distribution was thus the product of the mean and the coefficient of variation. The width of uniform distributions was twice the coefficient of variation (cv), so the minimum and maximum were mean ± 2 * cv * mean.

The reliability for total nitrogen was defined for being able to meet an effluent limit of 5 mg N L^{-1}, while that for ammonia was defined for being able to meet an effluent limit of 1 mg NH$_4$-N L^{-1}. 100,000 2-D (accounting for variability and uncertainty) Monte Carlo simulations were performed and the reliabilities for the two different effluent limits were recorded.

Sensitivity analysis was performed using Crystal Ball 5.0 after performing Monte Carlo simulations.

3. Results and Discussion

3a. Effects of Different Coefficients of Variation

Figure 1 shows that as the coefficient of variation increases, the reliability decreases, irrespective of the type of distribution used for the parameters. This is intuitively satisfying, as by definition an increased coefficient of variation increases the standard deviation. An increased standard deviation increases the likelihood of adverse combinations of variables leading to decreased reliability. In effect, as the combinations of parameters become less reliable, the simulated process should become less reliable. Figure 1 verifies this.

In this particular example shown in Figure 1, the design considered is for the 99th percentile flow and 99th percentile waste concentration. Other percentile flow and waste concentrations exhibited similar behavior.

Effects of Distribution Type

Figure 2 shows the effects of assumptions about the parameters' different distribution types on the reliabilities. This figure shows the reliability of each of the least-cost designs for different combinations of flow and influent waste concentrations. Reliability is defined in terms of meeting an effluent total N concentration of 5 mg N L^{-1}. 100,000 Monte Carlos simulations were performed.

The design assuming a lognormal distribution of parameters is 5-6% more reliable than the design assuming normal distribution. The design assuming a

2

Gaussian (or normal) distribution of parameters is 2-3% more reliable than the design assuming uniform distribution. Uniform distribution gives greater weight to the right tails of the distributions than either the lognormal or Gaussian distributions though less to the left tail of the lognormal distribution. As a result, a greater percentage of parameter combinations occur where adverse outcomes (effluent violations) result. Gaussian distribution has equally weighted left and right tails while lognormal has greater weight in the left than right tail though the right tail is longer. Because the reliability of a lognormal distribution results in a higher reliability than Gaussian distribution, one can conclude that increased values of the most sensitive parameters has a positive effect on meeting effluent limits.

3c. Effects of Correlated versus Uncorrelated Loadings

Figure 3 shows the effect of correlated versus uncorrelated loadings on the reliabilities of the various least-cost designs assuming the parameters have a Gaussian distribution. This figure shows the reliability of each of the least-cost designs for different combinations of flow and influent waste concentrations. In this particular figure, the coefficient of variation is 0.25. Reliability is defined in terms of meeting an effluent total N concentration of 5 mg N L^{-1}. 100,000 Monte Carlos simulations were performed.

While Figure 3 shows the reliability is approximately 1% greater for uncorrelated loadings than correlated for Gaussian distribution, the reliability is approximately 1.5% greater for uniform distribution and 3-4% greater for lognormal distribution.

3d. Robustness of Conclusion

Figures 4a, 4b, and 4c show the reliabilities in meeting ammonia and total nitrogen effluent limits for Gaussian, lognormal, and uniform distributions. In all cases, the coefficient of variation is 0.25 and the loadings are correlated. Reliability was defined as being able to meet a 1 mg NH_4-N L^{-1} for ammonia and 5 mg N L^{-1} for total nitrogen.

What is clear in all three figures (as well as Figures 2 and 3) is that spending additional money to meet higher flow and waste concentrations does not necessarily result in greater reliability of the process design. This is irrespective of the distribution type assumed for the parameters or whether the input loadings are correlated or uncorrelated (Figure 3). This is also irrespective of the coefficient of variation as shown in Figure 5. In Figure 5, the reliability of the least-cost designs for different flow and waste concentrations is shown assuming the parameters have lognormal distribution. This figure shows that the cost of the 80/80 (80th percentile flow, 80th percentile waste concentration) is less than the 90/80 design, but is more reliable.

4. Conclusions

In this paper it has been shown that the projected reliability of a wastewater treatment process to meet effluent ammonia (as N) and total nitrogen does depend upon the coefficient of variation assumed, the distribution type of the parameters assumed, and whether the input loadings are correlated or uncorrelated. The coefficient of variation assumed has the largest impact on the reliability values projected. These results are important to those making projections on the benefits of point source pollution prevention and those interested in permit trading.

These results also give further evidence that relying upon a least-cost design at particular flow and waste strength percentiles is not a particularly good means of design, for it is quite possible that a less expensive, more reliable design is possible.

Additional work of the effects of time series analysis would also seem to be a possible means of improving the estimations of reliability. This will be pursued in the near future.

Tables

Loading or Parameter	Distribution	Statistics (with units)
Flow	Lognormal	Mean = 47997; St. Dev = 13,285 ($m^3 \, day^{-1}$)
Temperature	Beta	Minimum=13.14, Maximum=18.13, α = 0.958, β = 0.864 (ºC)
COD	Logistic	Mean = 456.3; Shape = 70.62 ($mg \, L^{-1}$)
Nitrogen	Logistic	Mean = 22.8; Shape = 3.22 ($mg \, L^{-1}$)
Phosphorus	Logistic	Mean = 3.58; Shape = 0.68 ($mg \, L^{-1}$)
Suspended Solids	Logistic	Mean = 188; Shape = 45.0 ($mg \, L^{-1}$)
Ammonia Fraction	Uniform	Minimum = 0.6; Maximum = 0.8
Unbiodegradable Particulate COD Fraction	Uniform	Minimum = 0.07; Maximum = 0.20
Unbiodegradable Soluble COD Fraction	Uniform	Minimum = 0.04; Maximum = 0.10
Unbiodegradable Soluble TKN Fraction	Uniform	Minimum = 0.00; Maximum = 0.04
COD:VSS Ratio	G,L,U	Mean = 1.48
Endogenous Residue Constant	G,L,U	Mean = 0.2
Heterotrophic Yield	G,L,U	Mean = 0.45
MLVSS:MLSS Ratio	G,L,U	Mean = 0.75
N:COD Ratio	G,L,U	Mean = 0.1
Oxygen A Recycle Concentration	G,L,U	Mean = 1.5 mg O_2 L^{-1}
Oxygen S Recycle Concentration	G,L,U	Mean = 0.5 mg O_2 L^{-1}
Primary Specific Denitrification Rate	G,L,U	Mean = 0.224 day^{-1}
Secondary Specific Denitrification Rate	G,L,U	Mean = 0.100 day^{-1}
Readily Biodegradable COD Fraction	G,L,U	Mean = 0.165
Heterotrophic Decay Rate	G,L,U	Mean = 0.24 * (1.029 ** (temp – 20)) (day^{-1})
Nitrifier Decay Rate	G,L,U	Mean = 0.04 * (1.029 ** (temp – 20)) (day^{-1})
Nitrifier Growth Rate	G,L,U	Mean = 0.36 * (1.123 ** (temp – 20)) (day^{-1})
Nitrogen Saturation Rate	G,L,U	Mean = 1.0 * (1.123 ** (temp – 20)) (day^{-1})
Organic Nitrogen Conversion Rate	G,L,U	Mean = 0.015 * (1.029 ** (temp – 20)) (day^{-1})

Table a. Loading and Parameter Statistical Assumptions. (G=Gaussian, L=lognormal; U=uniform).

Flow %ile	80th	90th	90th	95th	95th	95th	99th	99th	99th	99th
Waste Concentration %ile	80th	80th	90th	80th	90th	95th	80th	90th	95th	99th
"s" recycle	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
"a" recycle	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
smf_{anoxic}	0.4042	0.411	0.4146	0.4092	0.4274	0.4277	0.4296	0.4246	0.4589	0.4270
$smf_{anaerobic}$	0.0528	0.0307	0.0563	0.0478	0.0297	0.0559	0.0275	0.0325	0.0411	0.0439
Reactor volume (m^3)	48,651	56,231	64,659	69,091	74,472	85,395	81,304	87,634	106,454	119,728
Total cost ($1000)	2,008	2,327	2,506	2,747	2,894	3,119	3,179	3,350	3,710	4,153

Table b. Least-cost Reactor Designs and Costs for Different Flow/Waste Concentration Combinations

Figures
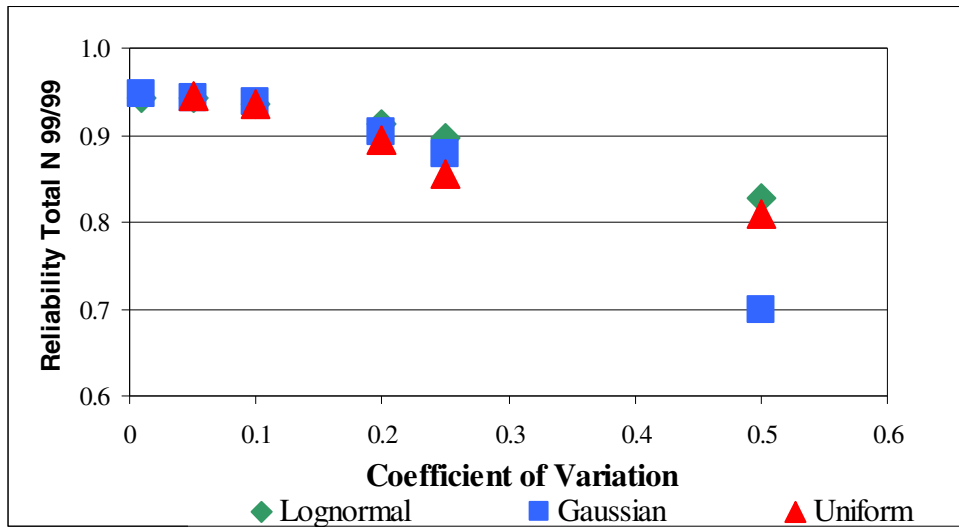
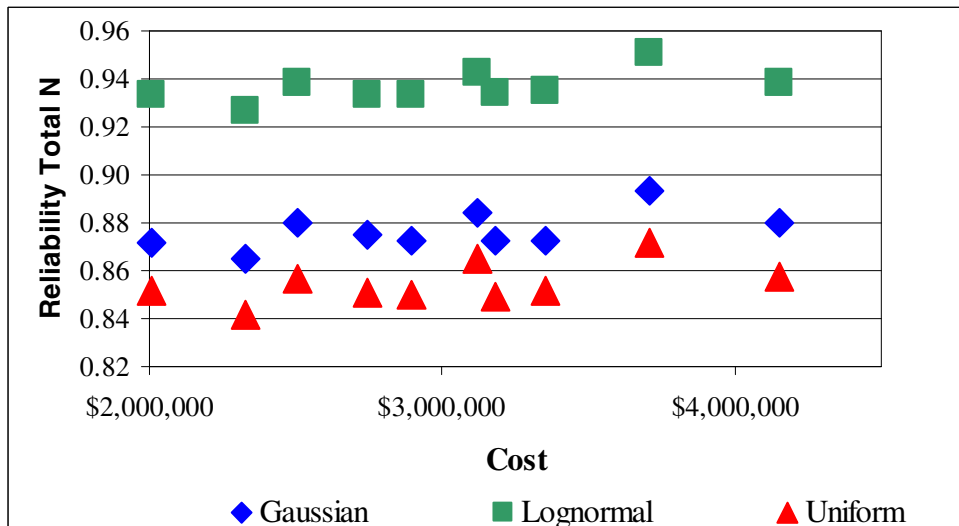Figure 1. Effect of Coefficient of Variation on Reliability of Meeting Total N Effluent Limit (Limit = 5 mg N L^{-1})

Figure 2. Effect of Different Distribution Types on Reliability for Meeting Effluent Total N Limit (Limit = 5 mg N L^{-1}, coefficient of variation = 0.25)
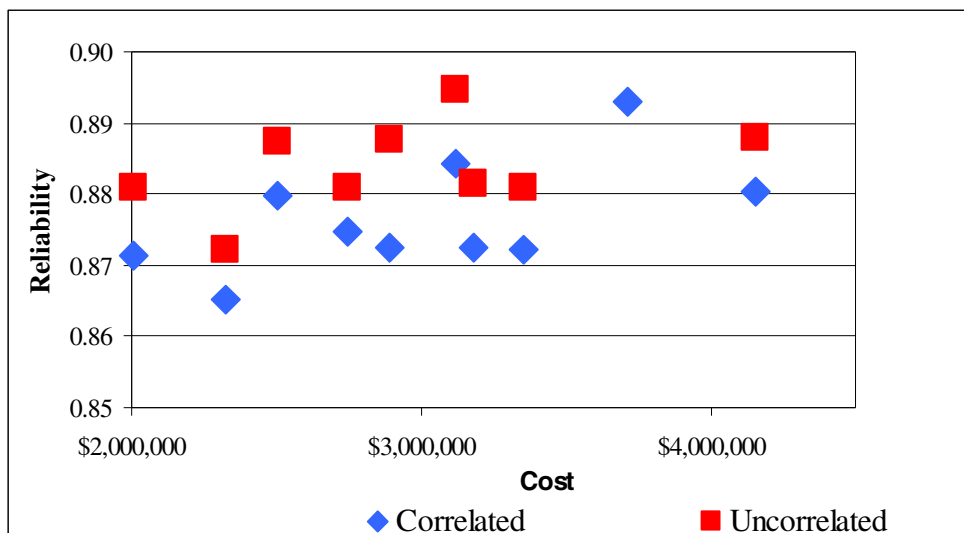
Figure 3. Effects of Correlated versus Uncorrelated Loading Values on Reliability for Meeting Effluent Total N Limit (Limit = 5 mg N L^{-1}, parameter distribution = Gaussian, coefficient of variation = 0.25)
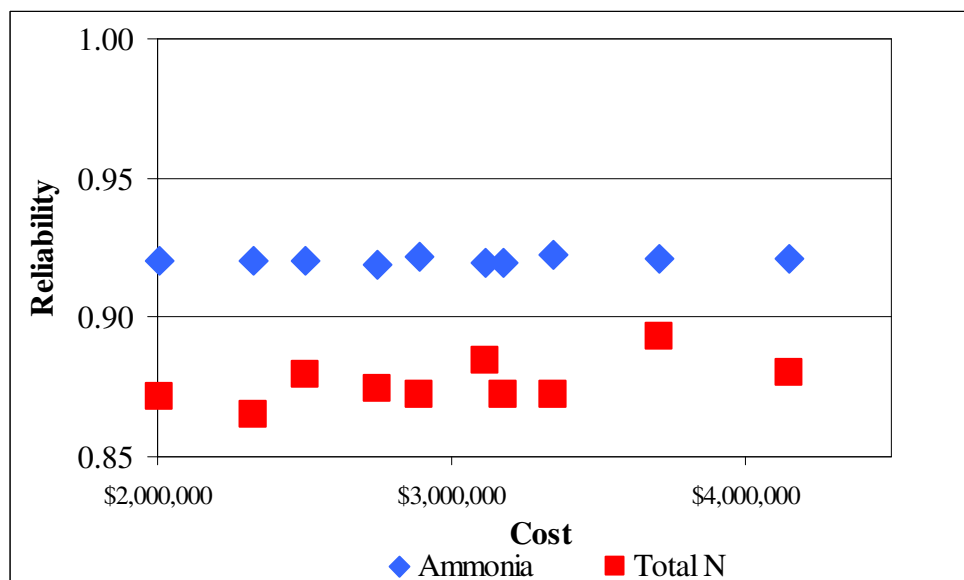
Figure 4a. Gaussian distribution: Cost versus Reliability for meeting effluent limits (1 mg NH_4-N L^{-1}, 5 mg total N L^{-1}) for different least-cost designs at different combinations of flows and waste concentrations. (Coefficient of variation = 0.25, correlated loadings)
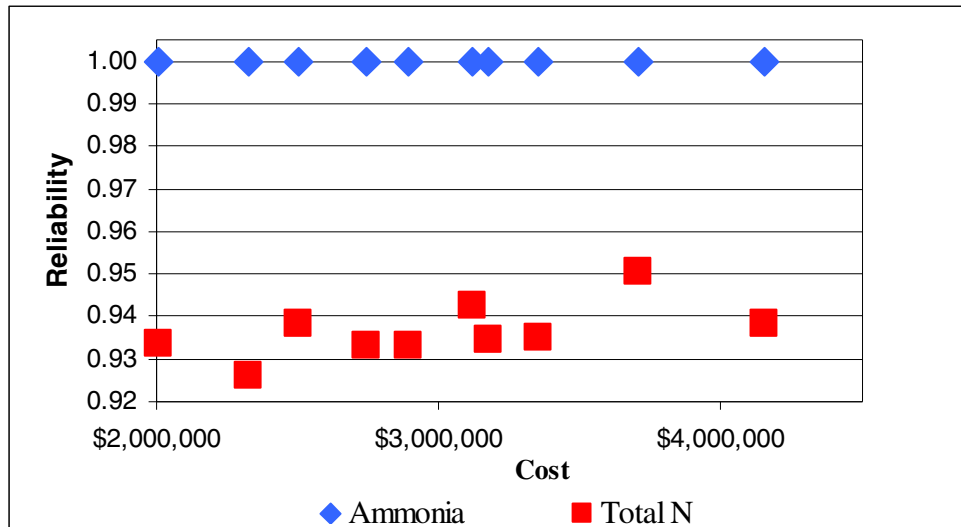
Figure 4b. Lognormal distribution: Cost versus Reliability for meeting effluent limits (1 mg NH_4-N L^{-1}, 5 mg total N L^{-1}) for different least-cost designs at different combinations of flows and waste concentrations. (Coefficient of variation = 0.25, correlated loadings)
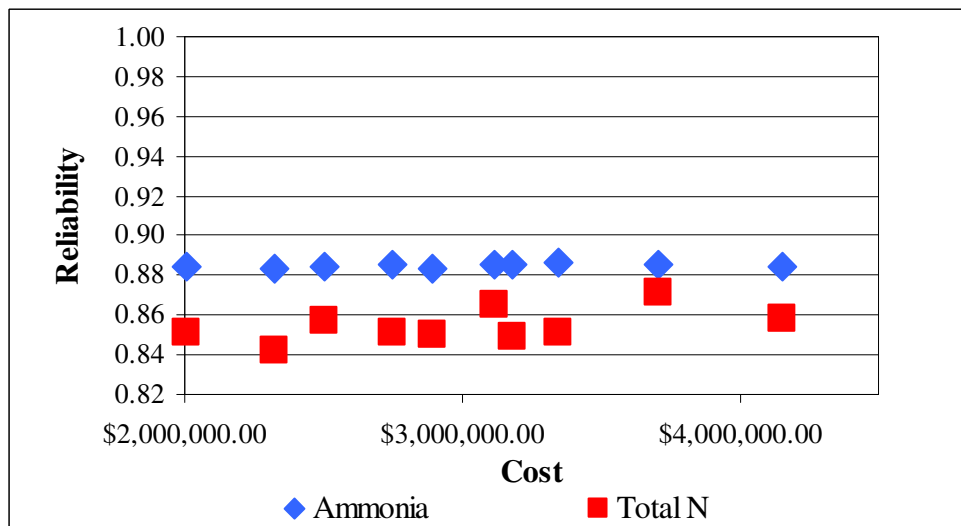

Figure 4c. Uniform distribution: Cost versus Reliability for meeting effluent limits (1 mg NH_4-N L^{-1}, 5 mg total N L^{-1}) for different least-cost designs at different combinations of flows and waste concentrations. (Coefficient of variation = 0.25, correlated loadings)
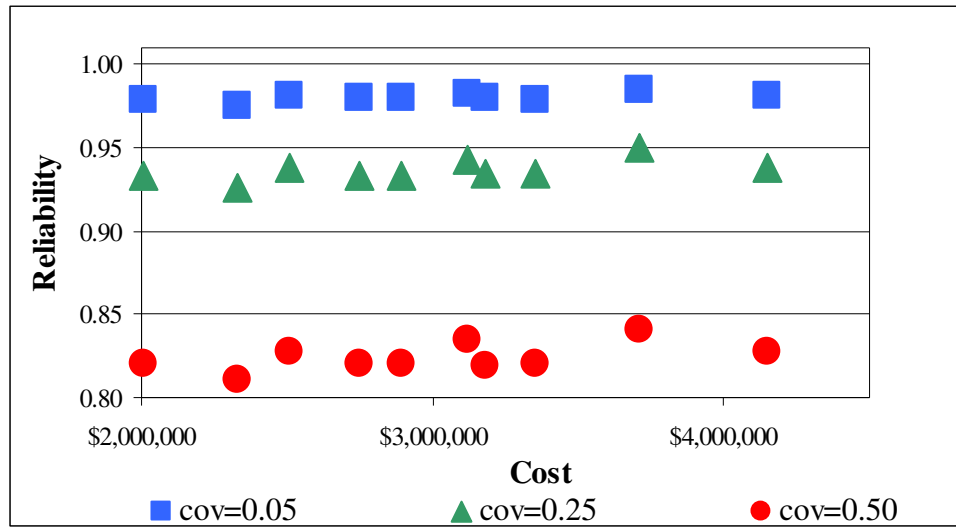
Figure 5. Reliability of meeting 5 mg total N effluent limit for least-cost designs assuming lognormal distribution of parameters for different coefficients of variability (cov)

References

Doby, T. A., De los Reyes, F. L., Baugh, J. W., Ducoste, J. J., and Loughlin, D. H. (submitted). "Assessing the reliability of deterministic least-cost wastewater treatment plant designs." *Journal of Environmental Engineering-ASCE*.

WRC. (1984). *Theory, Design and Operation of Nutrient Removal Activated Sludge Processes*, Pretoria, South Africa.